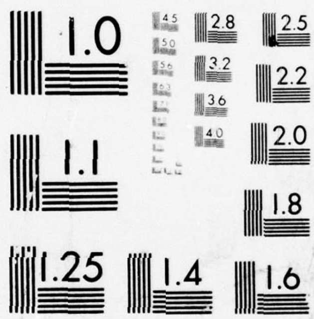
AD-A051 784 PURDUE UNIV LAFAYETTE IND DEPT OF STATISTICS F/G 12/1
LOCALLY OPTIMAL SUBSET SELECTION PROCEDURES BASED ON RANKS.(U)
SEP 77 S S GUPTA, D HUANG N00014-75-C-0455

UNCLASSIFIED MIMEOGRAPH SER-513 NL

| OF |

AD
A051 784

END
DATE
FILMED
4-78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(1) NW

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

(1)

Locally Optimal Subset Selection Procedures Based on Ranks*

by

Shanti S. Gupta, Purdue University

and

Deng-Yuan Huang, Academia Sinica, Taipei, Taiwan

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #513

December 1977

*This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Locally Optimal Subset Selection Procedures Based on Ranks*

by

Shanti S. Gupta, Purdue University

and

Deng-Yuan Huang, Academia Sinica, Taipei, Taiwan

In practice, it sometimes happens that the actual values of a random variable can only be observed under great cost or not at all, while their ordering is readily observable. This occurs for instance in life-testing when one only observes the order in which the parts under investigation fail without being able to record the actual times of the failure. Problems of this type suggest the investigation of decision rules based on ranks. Although the distributions of rank statistics are usually very involved, the resulting rules are often simple. Another advantage of rank procedures is that under the hypothesis that all distributions are identical, the distribution of the ranks does not depend on the underlying distribution. For this reason rank procedures are sometimes referred to as nonparametric rules. Hajek and Sidak [3] and others have developed an elegant theory of rank tests. Contributions to some related problems have also been made by Puri and Sen [6]. However, very little work has been done for multiple decision problems based on ranks. Gupta and McDonald [1], McDonald [4] and Nagel [5] have investigated several subset selection rules based on ranks. Nagel [5]

---

*This work was supported by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University.

tried to obtain some locally optimal rules based on the rank test theory. We are interested in deriving locally optimal ranking and selection procedures. Although, the criteria of optimality is different from Nagel's idea, the form of the procedures is the same. It has been shown that procedures of this type have many good properties.

From each of the k independent populations a fixed number of observations, say, n is taken. The distribution is assumed to depend on a parameter θ and the form of the distribution is also assumed to be known. Concluding that θ_i, $i=1,\ldots,k$, are or are not equal may not be sufficient. Often the experimenter is interested in ascertaining which population is associated with the largest (or smallest) θ, which populations possess the t largest (or smallest) θ, etc. Suppose the experimenter is interested in identifying which one of the k populations possesses the largest θ, the so-called "best" population. The subset selection approach to this problem is to select a small subset which is garanteed to contain the best population with probability P*, the basic probability requirement in these procedures. The selection of a subset including the best population is called a correct selection (CS). In this paper, we are interested in deriving procedures which satisfy the basic P*-condition and locally maximize the probability of a correct selection. An example is given to illustrate the application to a problem in regression analysis.

From each of the populations Π_i, $i=1,2,\ldots,k$, we take n observations $X_{i1},\ldots,X_{in}$. Let R_{ij} denote the rank of X_{ij} in the pooled sample of the $N=kn$ observations $(X_{11},\ldots,X_{1n};X_{21},\ldots,X_{2n};\ldots;X_{k1},\ldots,X_{kn})$.

We use the following definitions of Nagel [5]:

Definition 1. A rank configuration is an N-tuple $\Delta=(\Delta_1,\ldots,\Delta_N)$, $\Delta_i \in \{1,2,\ldots,k\}$, where $\Delta_i=j$ indicates that the ith smallest observation in the pooled sample comes from Π_j i.e. there exists an ℓ such that $R_{j\ell}=i$ holds.

Let $\mathscr{L}=\{\Delta\}$ denote the set of all rank configurations for the pair k and n which are kept fixed in these considerations. $\Delta_{\underline{x}}$ denotes the rank configuration of $\underline{x}=(x_1,\ldots,x_N)$. For a fixed Δ let $\mathscr{X}_\Delta=\{\underline{x}\in\mathscr{X}\mid\Delta_{\underline{x}}=\Delta\}$, where $\mathscr{X}=\{\underline{x}:\ \underline{x}=(x_1,\ldots,x_N)\}$. The decision space $\mathscr{D}$ consists of the 2^k-1 nonempty subsets d of the set {1,2, ...,k} and the empty set:

$$\mathscr{D}=\{d\mid d\subseteq\{1,2,\ldots,k\}\}.$$

A decision is the selection of a subset of the k populations. The fact that $i\in d$ indicates that Π_i is included in the selected subset if decision d is made.

Definition 2. A rank selection rule is a measurable function δ defined on $\mathscr{L}\times\mathscr{D}$, provided that for each $\Delta\in\mathscr{L}$, (i) $\delta(\Delta,d)\geq 0$ and (ii) $\sum_{d\in\mathscr{D}}\delta(\Delta,d)=1$ hold.

Let $\delta(\Delta,d)$ denote the probability that the decision d is made if the rank configuration Δ is observed.

Definition 3. A subset selection rule R based on ranks is a measurable mapping from $\mathscr{L}$ into R^k,

$$R:\quad \Delta\to(p_1(\Delta),\ldots,p_k(\Delta))$$

where $p_i(\Delta)=\sum_{d\ni i}\delta(\Delta,d)$ (summation over all subsets containing i).

If the p_i's take on the values 0 and 1 only then

$$\delta(\Delta,d)=\begin{cases}1 & \text{if } d=\{i|p_i(\Delta)=1\},\\ 0 & \text{otherwise,}\end{cases}$$

i.e. a non-randomized procedure is completely determined by its individual selection probabilities. Nagel [5] has shown that this is not true in general.

Let the distribution of Π_i be given by a density function $f(x,\theta_i)$ from a one-parametric family with the θ_i's belonging to some interval , which, without loss of generality, can be assumed to contain 0. Let $\Omega = \{\underline{\theta}|\underline{\theta} = (\theta_1,\ldots,\theta_k)\}$. Furthermore, let the family $f(x,\theta)$ have the following properties:

Condition A. (i) $f(x,\theta)$ is absolutely continuous in θ for almost every x;

(ii) the limit

(1) $\dot{f}(x,0)=\lim\limits_{\theta\to 0} \frac{1}{\theta} [f(x,\theta) - f(x,0)]$ exists for almost every x;

(iii)

(2) $\lim\limits_{\theta\to 0} \int_{-\infty}^{\infty} |\dot{f}(x,\theta)|dx=\int_{-\infty}^{\infty} |\dot{f}(x,0)|dx < \infty$

holds, with $\dot{f}(x,\theta)$ denoting the partial derivative with respect to θ. Note that the existence of $\dot{f}(x,\theta)$ for almost every θ is ensured at every point x such that $f(x,\theta)$ is absolutely continuous in θ. This, however, does not make the condition (ii) superfluous.

We know that if a density $f(x)$ is absolutely continuous and satisfies

$$\int_{\infty}^{\infty}|f'(x)|dx < \infty$$

then the family $f(x,\theta)=f(x-\theta)$ satisfies the conditions (i), (ii), and (iii) (see [5], p. 73), where $f'(x) = \frac{df(x)}{dx}$. And if a density $f(x)$ is absolutely

continuous and satisfies

$$\int_{-\infty}^{\infty} |xf'(x)|dx < \infty$$

then the family $f(x,\theta)=e^{-\theta}f[(x-\mu)e^{-\theta}]$ also satisfies the conditions (i), (ii), and (iii), (see [3], p. 73).

Our goal is to construct a selection rule δ based on ranks (δ is conditional on an observed rank configuration Δ) such that

(3) $\inf_{\underline{\theta}\in\Omega_0} P_{\underline{\theta}}(CS|\delta,\Delta) \geq P^*$ where $\Omega_0=\{\underline{\theta}:\ \theta_1=..=\theta_k\}$ holds and

(4) $P_{\underline{\theta}}(CS|\delta,\Delta)$ is as large as possible for all $\underline{\theta}$ in a neighborhood of $\underline{\theta}_0\varepsilon\Omega_0$.

Since in $\Omega_0=\{\underline{\theta}|\theta_1=\ldots=\theta_k\}$ the distribution of the ranks does not depend on the underlying distribution of the X_i's, $P_{\underline{\theta}}(CS|\delta,\Delta)$ is constant for $\underline{\theta}\varepsilon\Omega_0$. Hence, it suffices to choose any point in Ω_0 to satisfy (3). Hence without any loss of generality, we assume $\underline{\theta}_0 = (0,0,\ldots,0)$.

The probability that rank configuration Δ is observed under $\underline{\theta}$ with $\theta_i\neq 0$, $i=1,2,\ldots,k$, is

(5) $$P_{\underline{\theta}}(\Delta)=\int_{-\infty}^{\infty}\int_{-\infty}^{x_N}\ldots\int_{-\infty}^{x_2}\prod_{i=1}^{N} f(x_i,\theta_{\Delta_i})dx_1\ldots dx_N$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{x_N}\ldots\int_{-\infty}^{x_2}\prod_{i=1}^{N} f(x_i,0)dx_1,\ldots dx_N$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{x_N}\ldots\int_{-\infty}^{x_2}\left[\prod_{i=1}^{N} f(x_i,\theta_{\Delta_i})-\prod_{i=1}^{N} f(x_i,0)\right] dx_1\ldots dx_N$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{x_N}\ldots\int_{-\infty}^{x_2}\prod_{i=1}^{N} f(x_i,0)\, dx_1\ldots dx_N$$

$$+\sum_{i=1}^{k}\theta_i\sum_{\substack{j\\ \Delta_j=i}}\int_{-\infty}^{\infty}\int_{-\infty}^{x_N}\ldots\int_{-\infty}^{x_2}\frac{f(x_j,\theta_i)-f(x_j,0)}{\theta_i}\prod_{e=1}^{j-1} f(x_e,0)\prod_{e=j+1}^{N} f(x_e,\theta_{\Delta_e})dx_1\ldots dx_N$$

$$= A_0+\sum_{i=1}^{k}\theta_i A_i(\Delta,\underline{\theta}), \text{ where}$$

$$\prod_{i=1}^{m} f(x_i,0)=1, \text{ for } m=0,$$

$$A_0=\int_{-\infty}^{\infty}\int_{-\infty}^{x_N}\cdots\int_{-\infty}^{x_2}\prod_{i=1}^{N} f(x_i,0)dx_1\ldots dx_N$$

and

$$A_i(\Delta,\underline{\theta})=\sum_{\substack{j\\ \Delta_j=i}}\int_{-\infty}^{\infty}\int_{-\infty}^{x_N}\cdots\int_{-\infty}^{x_2}\frac{f(x_j,\theta_i)-f(x_j,0)}{\theta_i}\prod_{e=1}^{j-1} f(x_e,0)\prod_{e=j+1}^{N} f(x_e,\theta_{\Delta_e})$$

$dx_1\ldots dx_N$, $1 \le i \le k$.

Let G denote the group of permutations g of the integers $1,2,\ldots,k$:

(6) $$g(1,2,\ldots,k)=(g1,g2,\ldots,gk).$$

Let h be the inverse permutation of g, $h=g^{-1}$, and define

(7) $$g(x_1,\ldots,x_k)=(x_{h1},\ldots,x_{hk})$$

and for $d \in \mathcal{D}$, $gd=\{i|hi\varepsilon d\}$. Also for any $\Delta \in \mathcal{C}$, let $\bar{g}$ be defined as follows:

(8) $$\bar{g}\Delta=(g\Delta_1,\ldots,g\Delta_N),$$

$\bar{g}$ is thus induced by g. Let $\bar{G}$ be the group $\{\bar{g}\}$. And let $G(i,j)$ be the following subset of G

(9) $$G(i,j)=\{g\varepsilon G|gi=j\}.$$

<u>Definition 4.</u> A selection rule δ is invariant under permutation if and only if

(10) $$\delta(\bar{g}\Delta,gd)=\delta(\Delta,d) \text{ for all } \Delta\varepsilon\ \mathcal{C},\ d\varepsilon\ \mathcal{D},\ g\varepsilon G,\ \bar{g}\varepsilon\bar{G}.$$

Assume that Π_k is the best population then

(11) $$P_{\underline{\theta}}(CS|\delta,\Delta)=E_{\underline{\theta}}\sum_{\substack{d\\ d\ni k}}\delta(\Delta,d) = E_{\underline{\theta}}p_k(\Delta).$$

From the modified definition (11), it follows that a subset selection rule R is invariant under permutation if and only if

(12) $$(p_1(\bar{g}\Delta),\ldots,p_k(\bar{g}\Delta)) = g(p_1(\Delta),\ldots,p_k(\Delta))$$

for all $\Delta\varepsilon\, C$, $g\varepsilon G$, $\bar{g}\varepsilon\bar{G}$.

By (11) it is clear that an invariant rule R is completely defined by one of its individual selection probability functions $p_k(\Delta)$.

For invariant rules the probability of a correct selection can be expressed as follows:

$$E_{\underline{\theta}} p_k(\Delta) = \frac{1}{(k-1)!} \sum_{g\varepsilon G(k,k)} E_{g\underline{\theta}}\, p_k(\bar{g}\Delta) = \frac{1}{(k-1)!} \sum_{g\varepsilon G(k,k)} E_{g\underline{\theta}} p_{hk}(\Delta)$$

$$= \frac{1}{(k-1)!} \sum_{g\varepsilon G(k,k)} E_{g\underline{\theta}}\, p_k(\Delta) = \frac{1}{(k-1)!} \sum_{g\varepsilon G(k,k)} P_{g\underline{\theta}}(\Delta)$$

$$= A_0 + \frac{1}{(k-1)!} \sum_{g\varepsilon G(k,k)} \sum_{i=1}^{k} \theta_{hi} A_i(\Delta, g\underline{\theta}).$$

Since for any i,

$$\lim_{\underline{\theta}\to\underline{\theta}_0} \sum_{\substack{j \\ \Delta_j = i}} \frac{f(x_j,\theta_i) - f(x_j,0)}{\theta_i} \prod_{e=1}^{j-1} f(x_e,\theta_{\Delta_e}) \prod_{e=j+1}^{N} f(x_e,\theta_{\Delta_e})$$

$$= \sum_{\substack{j \\ \Delta_j = i}} \dot{f}(x_j,0) \sum_{\substack{e=1 \\ e\neq j}}^{N} f(x_e,0),$$

and for $\theta_i > 0$,

$$\sum_{\substack{j \\ \Delta_j = i}} \int\int \ldots \int \frac{|f(x_j,\theta_i) - f(x_j,0)|}{\theta_i} \prod_{e=1}^{j-1} f(x_e,0) \prod_{e=j+1}^{N} f(x_e,\theta_{\Delta_e}) dx_1 \ldots dx_N$$

$$= \sum_{\substack{j \\ \Delta_j = i}} \int \frac{|f(x_j,\theta_i) - f(x_j,0)|}{\theta_i} dx_j$$

$$= \sum_{\substack{j \\ \Delta_j = i}} \int \frac{1}{\theta_i} \left| \int_0^{\theta_i} \dot{f}(x_j,\theta)d\theta \right| dx_j$$

$$\leq \sum_{\substack{j \\ \Delta_j = i}} \int \frac{1}{\theta_i} \int_0^{\theta_i} |\dot{f}(x_j,\theta)| d\theta dx_j$$

$$= \sum_{\substack{j \\ \Delta_j = i}} \frac{1}{\theta_i} \int_0^{\theta_i} \int |\dot{f}(x_j,\theta)| dx_j d\theta,$$

and a similar result can be obtained for $\theta_i < 0$.

Hence

$$\limsup_{\underline{\theta}\to\underline{\theta}_0} \sum_{\substack{j \\ \Delta_j = i}} \int\cdots\int \frac{f(x_j,\theta_i)-f(x_j,0)}{\theta_i} \prod_{\ell=1}^{j-1} f(x_\ell,0) \prod_{\ell=j+1}^{N} f(x_\ell,\theta_{\Delta_\ell})dx_1\cdots dx_N$$

$$\leq \sum_{\substack{j \\ \Delta_j = i}} \int |\dot{f}(x_j,0)| dx_j .$$

By Dominated Convergence Theorem, we have

$$\lim_{\underline{\theta}\to\underline{\theta}_0} A_i(\Delta,\underline{\theta})$$

$$= \lim_{\underline{\theta}\to\underline{\theta}_0} \sum_{\substack{j \\ \Delta_j = i}} \int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} \frac{f(x_j,\theta_i)-f(x_j,0)}{\theta_i} \prod_{e=1}^{j-1} f(x_e,0) \prod_{e=j+1}^{N} f(x_e,\theta_{\Delta_e})dx_1\cdots dx_N$$

$$= \sum_{\substack{j \\ \Delta_j = i}} \int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} \dot{f}(x_j,0) \prod_{\substack{e=1 \\ e\neq j}}^{N} f(x_e,0)dx_1\cdots dx_N .$$

Now, there exists an $\varepsilon>0$ such that $0<|\theta_i|<\varepsilon$, for all i, $1\leq i\leq k$, $A_i(\Delta,\underline{\theta})$ is approximately equal to

$$\sum_{\substack{j \\ \Delta_j = i}} \int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{\infty}^{x_2} \dot{f}(x_j,0) \prod_{\substack{e=1 \\ e\neq j}}^{N} f(x_e,0)dx_1\cdots dx_N$$

$$= \sum_{\substack{j \\ \Delta_j = i}} B_j = A_i(\Delta),\ 1 \le i \le k, \text{ where}$$

$$B_j = \int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} f(x_j, 0) \prod_{\substack{e=1 \\ e \neq j}}^{N} f(x_e, 0)\, dx_1 \ldots dx_N.$$

We have

$$\sum_{g\varepsilon G(k,k)} \sum_{i=1}^{k} \theta_{hi} A_i(\Delta) = \sum_{i=1}^{k} \sum_{g\varepsilon G(k,k)} \theta_{hi} A_i(\Delta)$$

$$= (k-2)! \sum_{e=1}^{k-1} \theta_e \sum_{i=1}^{k-1} A_k(\Delta) + (k-1)!\, \theta_k A_k(\Delta)$$

$$= (k-2)!\ \{(U-\theta_k)V + (k\theta_k - U)A_k(\Delta)\}, \text{ where}$$

$$U = \sum_{i=1}^{k} \theta_i \text{ and } V = \sum_{i=1}^{k} A_i(\Delta) = \sum_{i=1}^{nk} B_i, \text{ independent of } \Delta.$$

V is zero if f is absolutely continuous and satisfies $\int_{-\infty}^{\infty} |f'(x)|dx < \infty$ ([3], p.66). Since $\theta_k \ge \theta_i$, $i=1,\ldots,k-1$, it follows that $k\theta_k - U > 0$. Hence $\sum_{g\varepsilon G(k,k)} \sum_{i=1}^{k} \theta_{hi} A_i(\Delta)$ is a nondecreasing function in $A_k(\Delta)$, thus we have proved the following result.

<u>Theorem.</u> If $f(x,\theta)$ satisfies the condition A, then for any i, $(1 \le i \le k)$,

$$p_i(\Delta)=\begin{cases} 1 & \text{if } A_i(\Delta) > c \\ \rho_i & \qquad = \\ 0 & \qquad <, \end{cases}$$

satisfies

$\inf_{\underline{\theta}\varepsilon\Omega_0} P_{\underline{\theta}}(CS|\delta,\Delta) \geq P^*$ such that $P_{\underline{\theta}}(CS|\delta,\Delta)$ is as large as possible in the neighborhood $0<|\theta_i|<\varepsilon, 1\leq i\leq k$, for given $\varepsilon > 0$. The constants ρ_i and c are determined by

$$\sum_{\substack{\Delta \\ A_i(\Delta) > c}} P_{\underline{\theta}_0}(\Delta) + \rho_i \sum_{\substack{\Delta \\ A_i(\Delta)=c}} P_{\underline{\theta}_0}(\Delta)=P^*. \tag{12}$$

Note that this locally optimal rule is based on weighted rank sums using the scores

$$B_i = \int_{-\infty}^{\infty} u^{i-1}(1-u)^{N-1}\varphi(u,f)du, \tag{13}$$

where

$$\varphi(u,f)= -\frac{\dot{f}(F^{-1}(u,0),0)}{f(F^{-1}(u,0),0)}. \tag{14}$$

Remark: (1) In problems concerning scale parameters, we use the condition $\int|xf'(x)|dx < \infty$ to replace $\int|f'(x)|dx < \infty$ to obtain $V = 0$.

(2) If the assumption $\underline{\theta}_0=(0,\dots,0)$ is replaced by the more general one $\underline{\theta}_0=(\theta,\dots,\theta)$,

$$\varphi(u,f,\theta)=\frac{\dot{f}(F^{-1}(u,\theta)\theta)}{f(F^{-1}(u,\theta),\theta)} \tag{15}$$

which in general depends on θ. However, it is independent of θ if θ is a location or scale parameter.

Nagel [5] has shown that the rules of this type are just provided that B_i's are non-decreasing in i, which for location parameters is

true if and only if $f(x)$ is strongly unimodal, i.e. if $-\log f(x)$ is a convex function ([3], p.20). It follows from Nagel [5] that $\inf_{\underline{\theta}\varepsilon\Omega} P_{\underline{\theta}}(CS|\delta,\Delta) = \inf_{\underline{\theta}\varepsilon\Omega_0} P_{\underline{\theta}}(CS|\delta,\Delta)$ for a just selection rule δ. If $f(x,\theta)$ is the normal density with mean θ and variance 1, then $\varphi(u,f) = \Phi^{-1}(u)$ where Φ is the cumulative distribution function of the standard normal random variable. Thus, the scores can be evaluated as

$$B_i = \int_0^1 u^{i-1}(1-u)^{N-i}\Phi^{-1}(u)du.$$

If f has the logistic density

$$f(x,\theta)=e^{-(x-\theta)}/\ [1+e^{-(x-\theta)}]^2$$

then $\varphi(u,f)=2u-1$ which leads to equally spaced scores: $B_i=a+ib$ where the actual values of a and $b > o$ are irrelevant. Hence the rule R_3 in [1] and [4],

R_3: Select Π_i iff $\sum_{j=1}^{n} R_{ij} \geq c$ is locally optimal on the respective P* level if the underlying distributions are logistic with location parameter θ. Nagel [5] has discussed a different type of optimality of R_3.

If $f(x_{ij},\theta_i)=f(x_{ij}-\theta_i\ C_{ij})$, $1 \leq i \leq k$, $j=1,\ldots,n$, the regression in location case, then we have for any i,

$$A_i(\Delta) = \sum_{\substack{j \\ \Delta_j=i}} C_{ij}B_j.$$

The procedure to select the population associated with the largest growth rate θ_i's is as follows: for any i,

$$P_i(\Delta) = \begin{cases} 1 & \text{if } A_i(\Delta) > c, \\ \rho_i & \quad\quad = , \\ 0 & \quad\quad < . \end{cases}$$

A related problem has been considered by Gupta and Huang [2] for the largest slope with unknown initial weight for nonparametric densities.

References

[1] Gupta, S. S. and McDonald, G. C. (1970). On some classes of selection procedures based on ranks. Nonparametric Techniques in Statistical Inference (ed. M. L. Puri), 491-514.

[2] Gupta, S. S. and Huang, D. Y. (1977). Nonparametric selection procedures for selecting the regression equations with largest slope. To appear in the journal of the Chinese Statistical Association, Taipei, Taiwan.

[3] Hajek, J. and Sidak, Z (1967). Theory of Rank Tests Academic Press, New York.

[4] McDonald, G. C. (1969). On some distribution-free ranking and selection procedure. Dept. of Statist., Purdue University. Mimeo Ser. #174.

[5] Nagel, K. (1970). On subset selection rules with certain optimality properties. Dept. of Statist., Purdue University, Mimeo Ser. #222.

[6] Puri, M. and Sen, P. K. (1971). Nonparametric Methods in Multivarite Analysis John Wiley & Sons, Inc. New York.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series #513	2. GOVT ACCESSION NO. 14 Mimeograph Ser-513	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 Locally Optimal Subset Selection Procedures Based on Ranks.		5. TYPE OF REPORT & PERIOD COVERED 9 Technical rept.
		6. PERFORMING ORG. REPORT NUMBER Mimeo. Series #513
7. AUTHOR(s) 10 Shanti S. Gupta [redacted] D. Y. Huang		8. CONTRACT OR GRANT NUMBER(s) 15 N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-243
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE 11 Sep [redacted] 77
		13. NUMBER OF PAGES 14 12 16p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Invariant rules, Subset selection, rank procedures, locally optimal selection rules.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper deals with subset selection rules based on ranks in the pooled sample. The procedures satisfy the P*-condition and also locally maximize the probability of a correct selection. An application to a problem in regression analysis is provided.

291 730

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)